

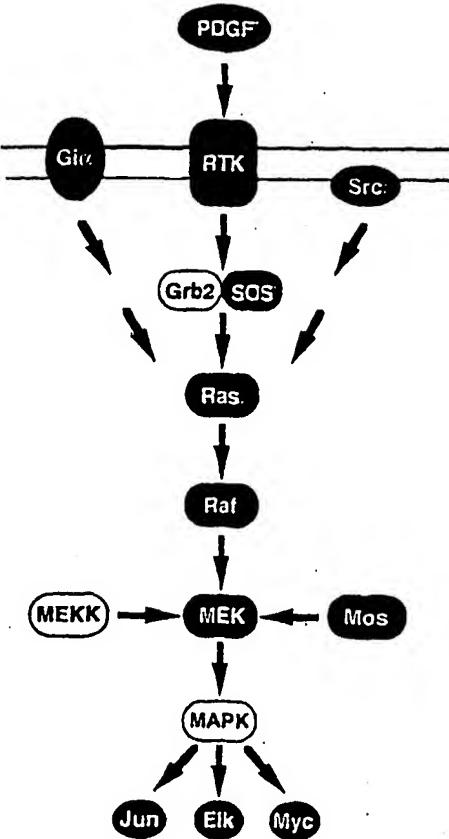
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY

(51) International Patent Classification <sup>7</sup> : <b>C12Q 1/68</b>		A1	(11) International Publication Number: <b>WO 00/52209</b> (43) International Publication Date: 8 September 2000 (08.09.00)
(21) International Application Number:	PCT/US00/05260		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(22) International Filing Date:	28 February 2000 (28.02.00)		
(30) Priority Data:	60/122,431	2 March 1999 (02.03.99)	US
(71) Applicant:	CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).		
(72) Inventors:	WILLIAMS, Lewis, T.; 125 Chapel Drive, Mill Valley, CA 94941 (US). RANDAZZO, Filippo; Apt. 403, 690 Chestnut Street, San Francisco, CA 94133 (US). FANTL, Wendy; Apt. 313W, 355 Buena Vista, San Francisco, CA 94117 (US).		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(74) Agents:	BLACKBURN, Robert, P.; Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608-2916 (US) et al.		

## (54) Title: MICROARRAYS FOR IDENTIFYING PATHWAY ACTIVATION OR INDUCTION

## (57) Abstract

The invention relates to a microarray containing polynucleotides representing genes encoding proteins of one or more biochemical pathways or cascades. The microarrays alternatively contain substrates for enzymes that are members of one or more biochemical pathways or cascades. The microarrays are used in methods of detecting changes in specific gene expression or protein function in relation to disease or treatment of disease.



***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## MICROARRAYS FOR IDENTIFYING PATHWAY ACTIVATION OR INDUCTION

## FIELD OF THE INVENTION

This invention concerns materials and methods relating to identifying genes that play a role in regulating pathways in cells. Aspects of the invention are exemplified in diagnosis of disease, characterization of biological conditions, gene discovery, genetic analysis, and treatment of disease.

## BACKGROUND OF THE INVENTION

Proteins that play a role in disease causation or prevention are often members of a group of proteins that make up a pathway or "cascade" of biochemical reactions in the normal or diseased cell, tissue, or blood. Examples of such pathways are the apoptosis pathway of cell death; the PI3 kinase pathway; the Wnt pathway; the Ras pathway; angiogenesis and anti-angiogenesis; and the tumor necrosis factor pathway.

It is possible to identify abnormal expression of individual members of these pathways and to correlate the expression with disease, such as cancer. However, it is more difficult to identify the point of regulation that has been disrupted, resulting in the disease state.

There is a need in the art for methods of detecting individual changes in gene expression and protein activity in a selected pathway or cascade in association with a disease state. There is also a need for identifying the key regulatory molecules that lead to the subsequent changes in the pathway or cascade. There is furthermore a need for a method of measuring the effect of a putative therapeutic agent on the identified regulatory molecule and the related pathway or cascade.

## 25 SUMMARY OF THE INVENTION

The invention provides a method of identifying regulatory genes for individual biochemical pathways of importance in disease such as cancer.

The invention also provides a method of detecting changes in expression of individual members of a biochemical pathway in a diseased cell or tissue in comparison with a normal cell or tissue.

5 The invention further provides a method of identifying genes, the expression of which is up-regulated or down-regulated by inducer molecules.

The invention also provides a method of identifying a pharmaceutical agent capable of altering expression of a regulatory gene, or capable of altering the activity of a regulatory gene product.

10 The invention provides a microarray having a plurality of polynucleotide molecules wherein the polynucleotide molecules represent genes encoding proteins that belong to a pathway sharing a regulatory gene or protein.

The invention also provides a microarray having a plurality of polynucleotide molecules, wherein the polynucleotide molecules represent genes each of which is up-regulated or down-regulated in a disease state such as cancer.

15 The invention further provides a microarray having a plurality of polynucleotide molecules, wherein a first group of the polynucleotide molecules represents genes of a first pathway, a second group of polynucleotide molecules represents genes of a second pathway, and an nth group of polynucleotide molecules represents an nth pathway; wherein each of the first, second or nth groups is applied to a 20 first, second, or nth sector of the surface of the microarray support, and wherein the number of sectors is n, with n being any integer from 2 to 1000.

25 The invention provides a method of generating a detectable pattern of gene expression for genes of one biochemical pathway wherein a microarray having a plurality of polynucleotide molecules representing genes encoding protein members of a biochemical pathway is exposed to nucleic acids obtained from a selected cell or tissue, and the pattern of hybridization of nucleic acids from the cell or tissue to the polynucleotides of the array indicates gene expression in the cell or tissue.

In a related embodiment, the invention provides a method of generating a detectable pattern of gene expression for genes of two or more biochemical pathways.

The invention also provides a method for detecting the effect of an inducer molecule or an inhibitory molecule on expression of one or more genes of a biochemical pathway, wherein a cell or tissue is exposed to the molecule, and the nucleic acids of the cell or tissue are exposed to a microarray having polynucleotides 5 representing genes of the pathway, and the hybridization pattern is detected and compared with the hybridization pattern of nucleic acid from a control cell or tissue.

The invention further provides a method for detecting the effect of an exogenous gene in a cell transduced with the gene, wherein the pattern of gene expression of the cell is detected by hybridization of nucleic acids from the cell to 10 polynucleotides of a microarray, wherein the polynucleotides represent genes encoding proteins of biochemical pathways, and the polynucleotides representing genes of each individual biochemical pathway are located in a defined sector of the microarray.

The invention also relates to a method of detecting the effect of a therapeutic agent administered to a mammal on the pattern of gene expression in 15 selected cells or tissues of the mammal, wherein the pattern of gene expression of the cell is detected by hybridization of nucleic acids from the cell to polynucleotides of a microarray, wherein the polynucleotides represent genes encoding proteins of biochemical pathways, and the polynucleotides corresponding to genes of each individual biochemical pathway are located in a defined sector of the microarray.

20 The invention still further relates to a microarray having a plurality of molecules capable of indicating enzymatic activity, wherein the molecules are arranged in sectors, wherein each sector contains molecules capable of indicating activity of an enzyme member of a selected biochemical pathway.

#### BRIEF DESCRIPTION OF DRAWINGS

25 Figure 1 shows the Ras signal transduction pathway: a cascade of oncogenes. Ras proteins function as critical intermediates between receptor (and nonreceptor) tyrosine kinases (RTKs) and downstream serine/threonine kinases (Raf-1, MEKs and MAP kinases). The Grb2/SOS complex links the activated receptor tyrosine kinase with Ras. A transforming version of the platelet-derived growth factor (PDGF)

was first identified as a potent retrovirus oncogene (v-sis). Mutated versions of the alpha subunit of the heterotrimeric G protein Gi have been identified in certain human tumors and can cause transformation via Ras-dependent signaling pathways. All components designated in black have been shown to exhibit transforming potential. To 5 date, no transforming versions of Grb2, MEK kinase (MEKK) or MAPKs have been identified.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

A goal of current biomedical research is to identify regulatory proteins and genes that ultimately determine the biochemical destiny of a cell, be it cell death 10 (apoptosis), differentiation, proliferation, quiescence, transformation to a tumor cell, metastasis to a remote site in the body, or participation in the normal function of a tissue through the production of specific proteins, antibodies, chemical messengers, and the like. Once these regulatory genes and proteins are identified, they become potential targets for intervention in the corresponding biochemical pathway.

15 Although the genes encoding proteins of important pathways continue to be identified, adequate information is known about them now to enable the design, production, and use of the microarrays of the invention. Specific pathways and embodiment are described herein, but the invention is not limited to these embodiments. The invention encompasses a method of applying related polynucleotides or proteins of 20 a pathway in a grouped or otherwise identifiable manner on a microarray, and of detecting changes in expression of the gene or activity of a related protein in a disease state, or as a result of exposure to a therapeutic agent, a regulatory molecule, or any other relevant perturbation of the cell, cell extract, tissue, organ, or whole organism.

25 For the biochemical pathways, the following are non-limiting embodiments of the microarrays:

##### 1. Single Pathway Microarrays

The microarray includes polynucleotides corresponding to genes encoding proteins of one biochemical pathway. A kit of this embodiment comprises a microarray, such as a glass slide or microchip, and optionally includes reagents for the

hybridization of probes to the array. This embodiment is suitable for screening cells, tissues, and any other biological sample to detect the expression of individual genes of this pathway. Optionally the kit contains on the same or on a separate microarray one or more polynucleotides corresponding to genes commonly referred to as 5 "housekeeping genes" as a control or an internal standard.

2. Multiple Pathway Microarrays

The microarray includes polynucleotides corresponding to genes of two or more pathways. Preferably each set of genes will be applied to a specific sector of the slide or microchip, with each sector containing at least 2 spots, preferably at least 4 10 spots, or more spots depending upon the number of different genes for that pathway. Each microarray can have at least two sectors, and the number of sectors can be increased one by one to n, where n is any integer from 2 to 1000. The number of sectors is limited only by the capacity of the microchip, the sphere, the microfluidic device, etc. A sector representing a specific pathway can appear once on the 15 microarray, twice, or multiple times.

A kit of this embodiment is prepared as described for embodiment (1) above, except that the microarray contains polynucleotides in sectors for the present embodiment, and the microarray is designed for hybridization of the sample material to two or more pathways in one step.

20 3. Microarray for Detecting Enzymatic Activity

The microarray can contain protein or polypeptide substrates for detecting the activity of an enzyme such as a kinase. As in embodiments (1) and (2), the microarray contains substrates of one pathway (embodiment 1), or two or more sectors (embodiment 2), each sector containing substrates for one pathway. This 25 microarray embodiment is designed to be exposed to a cell or tissue extract, a biological fluid, or a purified protein, for detecting enzymatic activity in the extract, fluid, or protein. If the enzymatic activity is phosphorylation, the substrate on the microarray will be capable of receiving a phosphate from ATP, or other suitable phosphate donor, if active kinase is present in the sample. A kit for this embodiment includes the 30 microarray, sphere, microfluid device or the like to which at least one set of substrates,

optionally at least two sectors of substrates, is attached, and optionally includes ATP or other suitable phosphate donor and a detectable signal system.

The following are non-limiting examples of embodiments of cell or tissue sources of nucleic acid probes and/or protein-containing cell extracts to which the 5 microarrays are exposed. All of these embodiments are suitable for the microarrays of the invention.

1. Nucleic acid probes or cell extracts are prepared from cell populations that have been transduced with a polynucleotide to determine the effect of the polynucleotide on expression of genes of one or more selected biochemical 10 pathways. The transduced cell can be a normal cell or a diseased cell such as a tumor cell. The transduced cell can be a cell from primary normal tissue, blood, primary tumor tissue, metastatic tumor tissue, or cell culture. The source of the transduced polynucleotide is not limited and can be all or part of a known gene, a polynucleotide corresponding to a known mRNA, or a polynucleotide of unknown function. In some 15 cases the polynucleotide is identified through its pattern of differential expression in one or more disease states. Examples of such polypeptides are disclosed in applications WO99/33982, WO99/38972, WO99/58675, and US99/22226. These applications are incorporated by reference herein.

2. Nucleic acid probes or cell extracts are prepared from cell 20 populations that have been treated with an inducer molecule, or inducer condition, defined herein as a molecule or condition that up or down-regulates gene expression or enzyme activity.

3. Nucleic acid probes or cell extracts are prepared from cells, tissues, or bodily fluids of an animal that has been exposed to a pharmaceutical agent, 25 an inducer molecule, or an inducer condition.

Inducer molecules and inducer conditions are not limited to those disclosed herein, and can be any materials or treatments that may affect gene expression or enzyme activity. Such inducer molecules and conditions include:

1. Molecules involved in proliferative or oncogenic pathways such as PI3 kinase; wnt; Ras, including k-Ras, h-Ras, and n-Ras; cABL; MEK; the family of FGFs; VEGF; KGF; IGFs; PDGF; and EGF;
2. Inducer Molecules that are involved in pro-angiogenic pathways, such as the family of FGFs, and VEGF;
3. Inducer Molecules that are involved in anti-angiogenic pathways, such as the 16K N-terminal fragment of prolactin;
4. Inducer Molecules that are small molecule antagonists or agonists of metabolic disease pathways, such as GSK 3 inhibitor that inhibits glycogen synthase kinase 3 and acts as an insulin-sensitizer, and a small molecule agonist of MC-4R that enhances cAMP production and that may be suitable for treatment of obesity;
5. Inducer Molecules that are involved in apoptotic pathways such as the fas Ligand, and TRAIL;
6. Inducer Molecules that are associated with cell adhesion and metastasis such as integrin;
7. Inducer Molecules can be cytokines and chemokines, such as IL-2, IL-12, SLC, BLC, and members of the TNF family;
8. Inducer Molecules can be interferons, such as IFN- $\alpha$ , - $\beta$ , and  $\gamma$ ,
9. Inducer Molecules can be any ligands, such as flt3, CD40, CD28;
10. Inducer Molecules can be any active fragments of the above (1-9), or any intracellular molecules that regulate the levels of RNA or DNA or other molecules;
11. Inducing Conditions can be osmotic shock or radiation such as UV or IR;
12. Other molecules can include molecules of the extracellular matrix.

This invention provides methods and materials for determining gene expression for genes of specific biochemical pathways, and kits including the necessary materials for practicing the method. The visualization of nucleic acid hybrids indicative

of gene expression, and the quantitations of the gene expression patterns, can be automated.

The method can be used for diagnosis of disease or diagnosis of any other biological or medical condition, for identifying new genes, for identifying 5 expression patterns in disease conditions, and for detecting changes in expression patterns following exposure to a pharmaceutical agent or procedure. The methods of the invention can be applied to any organism or to tissues or cells derived from any organism including animal or plant tissues or cells. In general, the method of the invention can be applied with equal utility to any living organism including but not 10 limited to vertebrates and invertebrates, such as mammals including humans, birds, fish, amphibians, reptiles, insects; eucaryotes and prokaryotes including fungi, bacterium and viruses; and all forms of plants.

The method of the invention employs probe populations of mRNA, or probe populations of cDNA derived by reverse transcription of mRNA, and also 15 employs preselected DNA. By the method of the invention, the probes may hybridize to preselected DNA molecules on the microarray, and the hybridization patterns are displayed.

In addition, for the purpose of normalizing the amount of probes in the populations of mRNA or cDNA being compared, a control gene or segment of 20 preselected DNA to which hybrids from both sources of probes are expected to form in equal amounts can be used. Such a gene can be, for example, any relatively invariantly expressed gene such as, but not limited to, for example, the actin gene.

In a preferred embodiment, the microarray consists of selected polynucleotides attached to a glass plate or wafer using methods known in the art. 25 Specific examples include the glass plate or wafer array methods as described in Wodicka et al., *Nature Biotechnology* 15:1359, 1997; glass microscope array methods as described in Schena et al., *Science* 270:467 (1995); and sialylated microscope slide methods of Schena et al., *Proc. Nat'l. Acad. Sci.* 93:10614 (1996).

The invention is not dependent on a specific array method, and other 30 methods can be used. These include three-dimensional arrays on spheres; microfluidic

devices employing beads; and non-fixed polynucleotides. Further, any suitable detection method can be used, such as a two-color detection method using fluorescein and lissanine (Schena et al., *Science*, supra); and fluorescent dyes Cy5 and Cy3 (Iyer et al., *Science* 283:83, 1999).

## 5 Exemplary Microarrays

### 4. Wnt, $\beta$ -caterin and APC

The Wnt signaling pathway is central to the development of some tumors, for example, Wnt is identified as a proto-oncogene in the mouse mammary gland (Tsukamoto et al., *Cell* 55:619, 1988). Wnt binds to its receptor and leads to 10 inactivation of GSK, glycogen synthase kinase. GSK normally phosphorylates  $\beta$ -caterin, which binds directly to the intracellular domain of E-cadherin, a protein involved in cell-cell adhesion. Inactivation of GSK causes hypophosphorylation of  $\beta$ -caterin, which in turn interferes with its binding to APC (adenomatous polyposis coli) protein. This  $\beta$ -caterin does not become ubiquitinated, and instead binds to Tcf/Lef (T 15 cell factor or leukocyte enhancing factor) and is translocated to the nucleus, where it binds to the promoter/enhancer region of one or more target genes and activates transcription. (Bullions et al., *Curr. Op. in Oncology*, 10:81, 1998.) Additional details of this pathway are known to those skilled in the art, and will provide guidance in selecting polynucleotides for a microarray. The overall pathway includes the following 20 steps that may be detected as different in normal and cancerous cells:

a. In some human tumors, levels of  $\beta$ -caterin are oncogenic. Screening of additional tumor samples and normal tissues using the microarrays of the invention will provide additional information relevant to regulation of  $\beta$ -caterin and to the design of therapeutic agents. Using a microarray having genes downstream from  $\beta$ -caterin, one of skill can determine the precise role played by  $\beta$ -caterin in tumor 25 development. This pathway also provides several substrates for a microarray useful for measuring enzymatic activity: phosphorylation of  $\beta$ -caterin; kinase activity of GSK; phosphorylation of APC; and ubiquitination of  $\beta$ -caterin. Such arrays can pinpoint the defect(s) in particular tumors, specifically, the ability of  $\beta$ -caterin to be phosphorylated

or ubiquitinylated; the ability of APC to be phosphorylated; and the enzymatic activity of GSK are a few examples. Such mechanisms are highly relevant to cancer, as truncations of APC, N-terminal deletions in  $\beta$ -caterin (deleting a GSK phosphorylation site), and point mutations of a GSK site in  $\beta$ -caterin have all been identified in human 5 tumors or cancer cell lines. Furthermore, the GSK consensus site is also required for ubiquitination and degradation of  $\beta$ -caterin. (Aberle et al., *EMBOJ* 16:3797, 1997.)

##### 5. PI3 Kinase and PTEN

There is evidence that several glioblastoma multiform cell lines exhibit elevated PKB/AKT (endogenous protein kinase  $\beta$ ) activity. This activity was abolished 10 by pretreatment with an inhibitor of PI3-kinase (phosphatidylinositol 3-kinase). The products of PI3-kinase activity include 3' phosphorylated inositol phospholipids, and this pathway is implicated in the inhibition of apoptosis, which relates to resistance to radiation and chemotherapy. (Leibel et al., *J. Neurosurg.* 66:1, 1987.)

A further element of this pathway is PTEN, a tumor suppressor gene that 15 encodes a phosphatase. PTEN can dephosphorylate the 3' phosphate of PI (3,4,5) P3, one of the lipid products of PI3-kinase. (Maehama et al., *J. Biol. Chem.* 273:13375, 1998.)

Cell transfection experiments have shown that the target of PTEN is upstream of, or at the level of, PI3-kinase. Evidence suggests that PTEN controls the 20 PI3-kinase-PKB pathway through the regulation of phospholipid levels (Haas-Kogan et al., *Current Biology* 8:1196, 1998). PTEN mutations have been identified in 60-70% of glioblastomas, advanced prostate cancers, and several sporadic and familial malignancies (Smith et al., *Cur. Biol.* 8:R241, 1998).

A variety of mutations in this pathway have been identified, such as a 25 P1-dependent protein kinase 1 mutant that is less dependent on 3' phospholipids, and mutants of PTEN that are less able to reduce PKB/Akt activity. The microarrays of the invention are designed to further elucidate the roles played by the components of the pathway, and to identify tumor cells and tissues in which defective proteins or altered gene expression play a role. A large number of biological samples can be screened for, 30 for example, decreased PTEN dephosphorylation activity (due to decreased gene

expression or to a mutation in the protein), or for activation of PK $\beta$ /Akt activity by growth factors, by activated PI-3 kinase, or by PDK1.

There is a need to determine whether PTEN mutations play a wider role in human cancers than is currently recognized. A microarray designed to screen for 5 changes in this pathway not only enables such determinations to be made, but also provides a system for efficient and rapid screening of candidate pharmaceuticals to remedy defective PTEN expression or activity.

#### 6. Protein Kinase B Family

This family shares a relationship with the PI-3 kinase pathway, which 10 has implications in designing microarrays that can be used to detect overall changes in the pathways, or to detect changes in specific components of the pathways. AKT2 is a member of the protein kinase B (PKB) family and has been shown to have oncogenic activity. AKT 2 is activated by mitogenic growth factors, including EGF, IGFI, IGFII, BFGF, PDGF, and insulin. In human ovarian cancer cells, this activation is mediated 15 through PI3-kinase. Activated Ras and Src also induced AKT2 kinase activity through PI3-kinase. (Liu et al., *Cancer Res.* 58:2973, 1998).

An oncogene, *c-p3k*, which encodes a member of the catalytic subunit, p110, of PI3-kinase, can transform cultured chicken embryo fibroblasts (Chang et al., *Science* 276: 1848, 1997). AKT3 is a downstream target of PI3-kinase and thus acts 20 through this oncogenic protein to mediate mitogenic signals to promote cell proliferation and cellular transformation.

#### 7. Ets Transcription Factors; Ras Pathway

The Ets family of transcription factors includes nuclear phosphoproteins that are involved in cell proliferation, differentiation and oncogenic transformation. 25 The family contains proteins having a conserved DNA-binding domain that forms a highly conserved helix-turn-helix structural motif. Ets proteins are targets of the Ras-MAPK signaling pathway. To direct signals to specific target genes, Ets proteins interact with other transcription factors that promote the binding of Ets proteins to composite Ras-responsive elements. (Waslylyk et al., *Trends. Biochem. Sci.* 6:213, 30 1998.)

The *ras* gene is a proto-oncogene essential for the growth and differentiation of various types of cells. The *ras* gene product, Ras, is a GTP binding protein that controls signal transduction by GTP hydrolysis. Recent evidence suggests that pathways related to Ras signaling may be deregulated in breast cancer cells. The 5 components of the Ras signal transduction pathway include a number of proto-oncogene proteins that control signaling events important in cell growth and differentiation.

Although only about 5% of breast cancers have mutations in *ras* itself, aberrant function of Ras-related proteins may contribute to breast cancer development. 10 Thus, mutations in genes encoding proteins upstream or downstream of Ras, and changes in activity of one or more of these proteins, may play a role in cancer development. Identification of such defects in individual cases of cancer may aid in designing specific therapeutics, and will also help clinicians to develop a pattern of changes in this pathway across a spectrum of cancer cases. Additional information on 15 members of the Ras pathway is available in Marshall et al., *Ann. Oncol.* 6:63-7, 1995.

#### 8. FGF and VEGF

Fibroblast growth factor (FGF) and vascular endothelial growth factor (VEGF) are important growth factors that play a role in normal cell growth and differentiation and are also implicated in oncogenesis. At least 18 FGF genes have been 20 identified, with varying effects on specific cell types and at specific stages of differentiation. FGF8 is implicated in cellular transformation, and other members of the FGF family play roles in tumor development. (Hu et al., *Mol. Cell. Bio.* 18:6063, 1998.) VEGF plays a role in angiogenesis, and is also implicated in development of cancer.

25 The microarrays of the invention provide methods of identifying the downstream effects of FGF and VEGF on specific cells and tissues.

#### Additional Embodiments

Where a new polynucleotide is identified by the methods of the 30 invention, standard molecular biology techniques for probing an appropriate library and

cloning the gene that hybridizes to the probe are employed to identify the gene by sequence and for further studies of the gene.

Comparisons of gene expression between, for example, normal and tumor tissue samples can be accomplished by creating hybridization patterns between 5 preselected DNA and two comparable populations of DNA or RNA probes from the tissues or cells being compared. The probes can represent transcripts from comparative populations of tissues or cells. The probes are differentially stained or labeled in order to make a visual comparison of the resulting hybridizations.

The information that can be derived by quantitating differential gene 10 expression from comparable tissues includes the expression patterns of genes from a given loci, the up or down regulation of certain genes in a certain biological condition, the location of particular gene or mutation of a gene on a chromosome or loci, any of which information can also be used to clone that gene if the gene is previously unknown.

15 The probes can be mRNA, or cDNA derived from mRNA, or a modified form of these. The modified form of cDNA can be, for example, derived from reverse transcription of an mRNA molecule using an oligo dT primer, in order to generate probes that contain at least the 3' sequence information of the gene, and which most likely include only the 3' end of the coding strand information on the cDNA first strand. 20 Thus, the cDNA can be a truncated first strand derived from the mRNA that can include sequence information that falls short of the entire coding region. Therefore, the truncated cDNA can contain, for example, from about 10 to about 20 nucleotides 5' of the oligo dT region (a sequence derived from the 3' end of the mRNA molecule) to about 400 or about 500 or more nucleotides 5' of the oligo dT region. The probe 25 population may contain any sequence length in between about 10 to about 500 nucleotides and may contain more than 500 nucleotides. A population of cDNA derived by reverse transcription of a population of mRNA molecules using an oligo dT primer will tend to be cDNA molecules favoring a representation of the 3' end of the mRNA from which they were transcribed, and these probes can range in length from

short to long, but most will contain at least a portion derived from the 3' most end of an mRNA molecule.

The two cell populations or tissue sources that are compared may be selected as directed by the nature of the comparison to be made for the differential 5 presentation of gene expression. Thus, for example, where one desires to compare gene expression in a type of cancer, mRNA or cDNA probes derived from a cancerous tissue can be compared to those derived from corresponding normal tissue, or probes derived from a population of immortal cells from a transformed cell line can be compared to the probes from normal, primary cells from a corresponding cell population. Where 10 another biological condition is the subject of the study, for example where stimulated tissue is to be compared to unstimulated tissue, the corresponding stimulated and unstimulated tissues or cell populations are selected as the sources from which the probes to be compared are derived. In general, for the purposes of this invention, stimulation can be any stimulation possible to apply to an animal, a tissue, or a cell, and 15 can include, for example, chemical, hormonal, physical or environmental stimulation. In general, any stimulation that may alter the gene expression pattern in comparable animals, tissues, or cells, is contemplated by the invention.

Preselected polynucleotides are components of the microarrays. The polynucleotides can be derived from normal tissue of normal animals, so that, for 20 example, a particular biological condition can be characterized in comparison to the hybrids that form between probes from normal tissue and normal genomic DNA, and hybrids that form between abnormal tissue and the normal genomic DNA. The preselected DNA can be any DNA, genomic or nongenomic, and will usually be a subportion of a genome of the animal being studied. Thus, for example, a chromosome, 25 or a portion of a chromosome, including, for example several loci may be selected as the preselected DNA and placed on the microscope slide. Although it is contemplated that the arrays for any specific pathway will contain polynucleotides corresponding to known genes of the pathway, the invention can also be used to determine if an unknown gene plays a role in a given pathway, by using polynucleotides of that gene as part of 30 the array for that pathway.

The ratio representing a differential gene expression in the two tissues that are being compared can range from 0 to infinity. The ratio can be 0, where, for example the hybridization pattern of the gene from one tissue is effectively in such excess to that in the comparative tissue that, for example, 0 hybrids are formed with 5 genes from the cancerous tissue, and, for example, 10 hybrids are formed with genes from the normal tissue, yielding a ratio of 0:10 or 0. This ratio indicates that the gene from the normal tissue is present in such excess in comparison to a homolog in the cancerous tissue that the gene from normal tissue successfully hybridizes to all the potential hybridization locations in the preselected polynucleotide on the slide. The 10 gene therefore may be, for example, a tumor suppressor gene.

At the other extreme is a ratio of infinity where, for example, a gene expressed in the cancerous tissue is in such excess to the homologous gene in the normal tissue, that the cancerous gene hybridizes to all the preselected DNA, no hybrids from the normal tissue are seen at that particular loci on the slide, and the ratio is 10:0. 15 This ratio indicates that the gene in the cancerous tissue gene is significantly up regulated as compared to that gene (or its homolog) in normal tissue, and thus may be an oncogene, and may be in part responsible for the cancerous condition.

Generally, the invention is practiced by creating populations of mRNA or cDNA probes from the tissues or cells, to be compared, normalizing the amount of 20 mRNA or cDNA probes against a gene that is invariantly expressed in these tissues, preparing preselected DNA, and affixing the preselected DNA onto a microscope slide by any method known in the art to achieve such effect, including but not limited to, for example, such methods described in Meng et al., *Nature Genetics* 9:432-438 (1995) and Cai et al., *PNAS USA* 92:5164-5158 (1995), differentially labeling or staining the two 25 populations of probes with two different optically identifiable stains or optically identifiable labels, hybridizing the two differently labeled populations onto the same microscope slide, viewing the differential staining pattern of the labeled hybrids that form with the preselected DNA on the slide, and quantitating the hybrids formed.

In addition, hybrids need to be normalized for their relative expression 30 levels by conducting a control hybridization to a preselected DNA representing a gene

the expression of which rarely varies from biological condition to biological condition.

The two populations of probes can then be adjusted for the amount of probe they contain relative to one another. To normalize the quantity of mRNA or cDNA probes from the two sources, a hybridization is conducted between the probes and preselected

5 DNA that is a gene the expression of which is relatively constant and predictable in all conditions, such as but not limited to, for example, the actin gene. This control is conducted using equal volume of the probe population from each source, and hybridizing the labeled probes to the same slide that has the actin gene affixed. The goal is to have the two populations hybridize in a ratio near 1:1, and any variation from  
10 this ideal ratio of 1 is used to normalize the rest of the hybridization experiments, by correcting for the variation in probe amount between the two populations of probes.

A nucleic acid molecule or nucleic acid sequence refers to either a DNA sequence, a RNA sequence, or complementary strands thereof that comprises a nucleotide sequence. The nucleic acid molecule can be a coding sequence, meaning

15 that the polynucleotide sequence of the nucleic acid molecule encodes a protein or portion or a protein. The nucleic acid molecule can also be a noncoding sequence such as an intron sequence, or any other polynucleotide sequence that is not expressed to form a functional protein. The term nucleic acid includes all known types of nucleic acid molecules, including polynucleotide sequences, messenger RNA or mRNA  
20 sequences, complementary DNA or cDNA sequences which are generally derived by reverse transcription of an mRNA sequence, total RNA such as, for example, that RNA that can be isolated from cells or tissues, and variations, modifications and derivatives of all these.

The term equal amount as used herein refers to approximately equal  
25 amounts of probes in the two probe populations being compared, and the amounts, where they are not exactly equal but only approximately equal can be normalized to correct for any lack of equality. For the purpose of normalizing the amount of probes in the populations of mRNA or cDNA being compared, a control gene or segment of preselected DNA that is relatively invariantly expressed, and to which hybrids from  
30 both sources of probes are expected to form in equal amounts can be used. Such a gene

can be, for example, any relatively invariantly expressed gene such as, but not limited to, the actin gene. The hybridization pattern of the two populations of probes to the normalizing or gene should, in the best possible circumstance be hybrids that form in a ratio of 1:1. However, where the ratio varies from a perfect ratio of 1, for example, 5 where the ratio of hybridization is 1:0.9 or 1.11, subsequent calculations to identify a ratio of the differential hybridizations that result with the preselected DNA that is the object of the genetic analysis being conducted can be corrected or normalized with the adjusted ratio of 1.11, to reflect the relative proportions of probe amount in the two populations. Therefore, the term "equal amount" refers to an approximately equal 10 amount, and in any event, to an amount that can be normalized or corrected for any variation from the ideal ratio of 1. The method of the invention, while quantitative in the sense that a quantifiable amount of hybridizations can be compared to reflect differential gene expression, does reflect an approximation, albeit a close approximation, of the relative hybridizations and so a close approximation of the 15 relative gene expression of the two populations of probes being compared. Where a thorough study of gene expression in a given population of probes is desired, it is implicit in the method of the invention that several repeated hybridizations with the same preselected DNA should be conducted so that an average representative ratio of the gene expression levels being studied can be attained.

20 The term preselected polynucleotides refers to polynucleotides that are affixed to the microarray and that will form hybrids with the mRNA or cDNA probes of the invention. The polynucleotide can be genomic DNA, including subsets of a genome. For example, the preselected polynucleotide can be several loci of a particular chromosome at which it is expected differential expression patterns in comparable cells 25 or tissues may be identified. The preselected polynucleotides can also be cDNA or in general any polynucleotide to which hybrids will form in contact with nucleic acid probes.

30 Examples of particular preselected polynucleotide molecules are polynucleotides derived from a vertebrate or invertebrate species, such as a mammalian gene, including a human gene, an avian gene, a fish gene, an amphibian gene, a reptilian

gene; polynucleotides derived from eucaryotes and prokaryotes, such as a fungal gene or a bacterial gene; polynucleotides derived from a virus; polynucleotides derived from an insect; and polynucleotides derived from a plant. In preferred embodiments of the invention, as described herein, the preselected polynucleotides represent genes encoding 5 proteins that are members of a biochemical pathway or cascade.

The term diseased as used herein refers to tissue from an organism that is afflicted with an adverse biological, physiological, genetic, or molecular biological condition. The tissue taken from the organism, and the cells that make up the tissue, can also be considered diseased, and can manifest the disease before the entire organism 10 is considered. The optimal function can be indicated by any number of indicia, including but not limited to an up or down regulation in gene expression. The disease can also be identified by a variety of other parameters, including but not limited to the condition of the organism generally, the condition of tissue taken from the organism as compared to corresponding normal tissue, and the condition of cells that make up tissue 15 taken from the organism. The disease can be an identified or as yet unidentified disease, and may be a disease that carries a variety of phenotypes, genotypes, or physiological or biological effects. The disease may be in the initial stages of the progression of the disease, or may in the later stages. The disease may be infectious or noninfectious.

20 The term inflamed as used herein refers to a condition of tissue in which the tissue condition indicates that the organism is experiencing a local immune reaction known as inflammation, or inflammatory response. An inflammatory reaction may occur because of any number of causes, including but not limited to infection, the presence of nonself antigen in the organism, an allergen, an irritation, chemical 25 treatment or poisoning, a therapy or treatment for another condition that induces the side effect of inflammation, a complement mediated reaction, an allergic reaction, a T-cell reaction, and an antibody reaction. The condition of inflammation in a tissue may be short lived, or chronic, and it is possible that the true causes of the inflammation are unknown. In any event a diagnosis of inflammation may be made in a variety of 30 circumstances by a variety of methods.

The term cytokines as used herein refers to a broad class of molecules that are capable of promoting cell growth and division. The class includes, but is not limited to the interleukins including IL-1, IL-2, IL-3, IL-6, IL-8, and IL-10, colony stimulating factors such as, for example GM-CSF, and interferon such as, for example, 5 IFN-1. Stimulation by cytokines can occur by inducing cytokine production endogenously in the cells, tissue, or organism, or by administering exogenous cytokines to the organism, tissue or cells.

The term growth factor as used herein refers to a member of a class of molecules that are unified by their ability to promote cell growth. Exemplary growth 10 factors include, but are not limited to, for example, platelet derived growth factor (PDGF), fibroblast growth factor (FGF), insulin-like growth factor (IGF), transforming growth factor (TGF), epithelial growth factor (EGF), and nerve growth factor (NGF). Growth factors have various mechanisms of action and various structures and are characterized only by the capability to promote cell growth. The invention also 15 embodies the use of growth factors that are not presently known, but which may be discovered at a later date.

The term differentiation factor as used herein refers to a molecule that can cause a cell or tissue to differentiate, i.e., to become more specialized in its function. Because differentiation is sometimes considered a process specific to a 20 particular pathway or to a particular final differentiated result, differentiation factors can be specific to tissues or cells in which the differentiation occurs.

The term physically stimulated as used herein refers to any physical stimulation that can be applied to an organism, a tissue, or a cell. Physical stimulation can include, but is not limited to, for example radiation, electrical, heat, cold, light, 25 rigor, starvation of any kind, environmental stimulus, the passage of time, or in general any physical stimulus that is believed might alter the gene expression of a given cell, tissue, or organism.

The term chemically stimulated as used herein refers to any chemical stimulation that can be applied or administered to an organism, a tissue, or a cell. 30 Chemical stimulation can include, but is not limited to, for example, administration of

toxins, drugs, small molecules, inhibitors, plants or chemicals derived from plants. Chemical stimulation may also include stimulation by more than one chemical, either, for example, by administering more than one chemical at the same time, or by administering the chemicals consecutively.

5        The term optically detectable marker as used herein refers to any of a number of optically detectable markers, labels or stains that can affix, bond, or otherwise identify a nucleic acid molecule, and which can allow visualization of the presence of the nucleic acid molecule to which the label is affixed. Several optically detectable markers are available from such companies as, for example, Stratagene, La  
10      Jolla, California, and Promega, Madison, Wisconsin. The label for labeling nucleic acid molecules can be any label that will visually identify the nucleic acid molecule, including, but not limited to, for example, fluorescein and rhodamine.

15      The term hybridization as used herein refers to the formation of a binding pair of molecules. The binding pair can be a pair of molecules including, for example, a DNA/DNA pair, or a DNA/RNA pair, in which the components of the pair bind specifically to each other with a higher affinity than to a random molecule. The pairing between nucleic acids molecules can be between two single stranded molecules, or a single and a double stranded molecule. An optically detectable label can be attached to one or both of the hybridizing molecules for visualization of the molecules.  
20      Hybridization can occur upon contact of the two molecules of the binding pair, which contact may occur, for example, on a microscope slide, a glass plate, in a tube, a petri dish, on plastic or other synthetic material, and in vitro or in vivo. The term "specific binding" used in reference to interaction between two molecules to form a binding pair indicates a higher affinity binding and a lower dissociation constant than nonspecific  
25      binding, thus, distinguishing specific binding from background binding. The hybridization conditions for forming the DNA/DNA hybrids or the RNA/DNA hybrids of the invention can be any hybridization condition that will allow a hybrid to form, and can include, but are not limited to those techniques described in Ausubel et al. (1994) Current Protocols In Molecular Biology, (Greene Publishing Associates and John Wiley

& Sons, New York, NY), and Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor Press, Cold Spring Harbor, New York).

The term kit as used herein refers to a package containing the specified material and includes printed instructions for use of the material. For example, a kit can 5 be a diagnostic kit for identifying the characteristics or progress of a disease by identifying up or down regulation of specific genes in certain diseases. Printed instructions may be written or printed on paper or other media, or committed to electronic media such as magnetic tape, computer-readable disks or tape, CD-ROM, and the like. Kits may also include plates, tubes, microscope slides, stain, dishes, diluents, 10 solvents, wash fluid or other conventional reagents, and may also include polynucleotide materials.

The method of the invention can be used to isolate new genes that are identified during probe hybridization with preselected polynucleotide to result in a gene expression pattern that highlights a particular gene or group of genes. The isolation of 15 new genes can be accomplished by identifying the preselected polynucleotide that forms the hybrid, isolating that piece of polynucleotide, cloning the polynucleotide sequence into a vector, and using the sequence to probe a cDNA library. The gene that hybridizes to preselected polynucleotide can then be cloned by standard molecular biology techniques. The information about any gene so identified can be used as a basis 20 for developing therapeutic strategies to fight the disease or biological condition in which the gene is up regulated or down regulated, or which is implicated in the mechanism of the disease or condition. The effectiveness of any such therapy can be determined by reapplication of differential presentation of gene expression using probes derived from the treated tissue as compared to probes from either the nontreated normal 25 tissue, or the nontreated diseased tissue. Such analysis of the effectiveness of any therapy or treatment can be applied using the method of the invention at any point during the course of a therapy or treatment.

The method of the invention may also be used to study differential expression of plant genes for various uses in understanding plant development, and 30 plant disease and for genetically engineering plants.

The tissues that can be studied by the method of the invention include any tissues from which nucleic acid (DNA or RNA) can be extracted, or from which cDNA can be generated. Thus, for example, where the method of the invention is used to diagnose or characterize a particular cancer, the nucleic acid can be generated from 5 the cancer tissue, for example, in the case of liver cancer, from liver tumor tissue, and can be compared with normal liver tissue. Similar comparisons can be made between normal and affected or afflicted tissues in all cases where the method of the invention is applied to treat or diagnose a disease or other biological condition. Comparisons can also be made between treated and untreated tissues, so that, for example, comparisons 10 can be made between the expression patterns of tissue stimulated with, for example, cytokines, growth factors, or differentiation factors, or the same tissues unstimulated. Other stimulus can include, for example, chemical or physical stimuli, or environmental stimuli.

The present invention will now be illustrated by reference to the 15 following examples which set forth particularly advantageous embodiments. However, it should be noted that these embodiments are illustrative and are not to be construed as restricting the invention in any way.

### EXAMPLES

20

#### EXAMPLE 1

##### MICROARRAY PREPARATION, HYBRIDIZATION, AND SCANNING

Microarray Preparation. Amino-modified PCR products are suspended at a concentration of 0.5 mg/ml in 3x standard saline citrate (SSC) and arrayed from 96-well micro-titer plates onto sialylated microscope slides (CEL Associates, Houston) 25 using high-speed robotics. cDNAs are arrayed in 1.0-cm<sup>2</sup> areas. Printed arrays are incubated for 4 hours in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 minute, twice in H<sub>2</sub>O for 1 minute, and once for 5 minutes in sodium borohydride solution (1.0 g of NaBH<sub>4</sub> dissolved in 300 ml of PBS

and 100 ml of 100% ethanol). The arrays are submerged in H<sub>2</sub>O for 2 minutes at 95°C, transferred quickly into 0.2% SDS for 1 minute, rinsed twice in H<sub>2</sub>O, air dried, and stored in the dark at 25°C.

Tissue mRNAs are purchased (CLONTECH). Cellular mRNA is  
5 isolated using methods as described by Schena et al., *Proc. Nat'l. Acad. Sci.* 93:10614  
(1996). Probes are made as described in Schena et al. with several modifications. The  
reverse transcriptase used is Superscript II RNase H (GIBCO). The Cy5-dCTP is  
purchased from Amersham. Each reverse transcription reaction contains 3.0 µg of total  
human mRNA. For quantitation, the mRNAs are doped into the reverse transcription  
10 reaction at ratios of 1:100,000, 1:10,000, and 1:1000 (wt/wt) respectively. Following  
the reverse transcription step, samples are treated with 2.5µl of 1 M sodium hydroxide  
for 10 minutes at 36°C, then neutralized by adding 2.5µl of 1 M Tris-HCl (pH 6.8) and  
2.0 µl of 1 M HCl. Probe mixtures contain cDNA products derived from 3 µg of total  
mRNA, suspended in 5.0 µl of hybridization buffer (5 x SSC plus 0.2% SDS).

15 Hybridization and Scanning. Probes are hybridized to 1.0-cm<sup>2</sup>  
microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C in a  
hybridization chamber as described in Schena et al. Arrays are washed for 5 minutes at  
room temperature (25°C) in low stringency wash buffer (1 x SSC/0.2% SDS), then for  
10 minutes at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS).  
20 Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device, fitted with  
a custom filter set (Chroma Technology, Brattleboro, VT). Accurate differential  
expression measurements (i.e., final fluorescence ratios) are obtained by taking the  
average of the ratios of two independent hybridizations.

## EXAMPLE 2

25

### PREPARATION OF cDNA CLONES

Human cDNA Clones. The cDNA library is made with mRNA from selected cells, for example cells transduced with polynucleotide sequence of unknown function. Inserts >600bp are cloned into the lambda vector λYES-R to generate 107-

108 recombinants. Bacterial transformants are obtained by infecting *E. coli* strain JM107/λKC. Colonies are picked at random and propagated in a 96-well format, and minilysate DNA is prepared by alkaline lysis using REAL preps (Qiagen, Chatsworth, CA). Inserts are amplified by PCR in a 96-well format using primers (PAN132, 5'-  
5 CCTCTATACTTTAACGRCAAGG; and PAN133, 5'-TTGTGTGGAATTGRGA-GCGG) complementary to the λYES polylinker and containing a six-carbon amino modification (Glen Research, Sterling, VA) on the 5' end. PCR products are purified in a 96-well format using QIAquick columns (Qiagen).

### EXAMPLE 3

10 PREPARATION OF A MICROARRAY FOR RAS-MAP-KINASE PATHWAY GENES IMPLICATED  
IN CELL PROLIFERATION AND ONCOGENESIS.

Using the methods described above, or any other appropriate methods for obtaining polynucleotides, applying them to microarrays, and detecting hybridization, a microarray for an Ets-related pathway is prepared using pathway genes  
15 as described in detail in Wasylyk et al., *Trends in Biochem. Sci.* 23:213 (1998). The Ets family of transcription factors includes proteins that serve as targets for Ras-induced phosphorylation. A suitable array will comprise polynucleotides of the following genes:

MAP kinase members ERK, JNK, and p38/RK  
20 Ets family members ETS, YAN, ELG, PEA3, ERF and TCF

This pathway is also suitable for preparing a microarray of enzyme substrates, specifically kinase substrates (Figure 1). As described by Wasylyk et al., phosphorylation of an Ets protein enhances its ability to activate transcription. This in turn is a product of upstream kinase activation of MAP kinase members that  
25 phosphorylate the Ets proteins.

An array of this pathway can elucidate changes in gene expression and protein phosphorylation that are brought about by putative tumorigenic stimuli such as

growth factors. The role of this pathway in individual cancers can be studied, as can the effect of an agent administered to inhibit the pathway.

#### EXAMPLE 4

5

#### GENE DISCOVERY USING A RAS PATHWAY MICROARRAY

Figure 1 illustrates in schematic form the *Ras* signal transduction pathway. It is believed that other members of the family of genes may be oncogenes. (Clark et al., *Breast Cancer Research and Treatment* 35:133, 1995.)

10

A microarray comprising genes upstream and downstream from ras, plus one or more genes of unknown function that are upregulated in tumor cells, can be used to identify whether the unknown genes are regulated by the same factors that regulate the ras pathway.

15

The foregoing are a few examples of individual pathways. As described herein, a microarray can comprise polynucleotides relating to one pathway, or sectors each containing genes relating to one pathway. Additional information for preparing such microarrays includes: Marshall et al., *Ann. Oncol.* 6:563, 1995 (ras pathway);

Vojtek et al., *J. Biol. Chem.* 273:19925, 1998 (ras pathway and related mutations);

20 Chiarugi et al., *Int. J. Mol. Med.* 2:715, 1998 (cyclooxygenase pathway; angiogenesis);

Shimizu et al., *Hum. Cell* 9:175, 1996 (PI3 kinase pathway; adhesion); Brown et al.,

*Curr. Opinion Cell Biol.* 10:182, 1998 (wnt signaling pathway); Alessi et al., *Biochem.*

*Biophys. Acta* 1436:151, 1998 (AKT pathway); Jankowski et al., *Mol. Pathol.* 50:289,

1997 (E-cadherin pathway); and Kumar et al., *Oncogene* 17:1365, 1998 (Protein Kinase

25  $\beta$ ).

Further information on pathways is available on the World Wide Web, for example the Kyoto Encyclopedia of Genes and Genomes provides information on the relationship of a gene or protein to a pathway.

## CLAIMS

1. A method for identifying differential gene expression in first and second mammalian cells, said method comprising

- (a) exposing said first cell to an agent capable of transforming said cell;
- (b) providing a second mammalian cell which is not exposed to said agent;
- (c) extracting mRNA from said first and second cells;
- (d) preparing first and second cDNA libraries from said mRNA;
- (e) contacting each of said cDNA libraries with a first and second microarray comprising a first set of polynucleotides each of which is at least 50 base pairs in length, and each of which has the partial sequence of a specific mRNA corresponding to a gene encoding a protein member of a first defined biochemical pathway;
- (f) detecting hybridization of said first cDNA library with said first microarray and said second cDNA library with said second microarray, wherein differential hybridization of said first cDNA compared to said second cDNA indicates differential expression of one or more genes of said pathway.

2. The method of claim 1, wherein said microarray further comprises a second set of polynucleotides having the partial sequence of mRNA corresponding to a gene encoding a protein member of a second defined biochemical pathway, wherein each of said first and second sets of polynucleotides is confined to a separate sector of said microarray.

3. The method of claim 2, wherein said microarray further comprises a plurality of sectors each comprising a set of polynucleotides corresponding to genes encoding different biochemical pathways.

4. The method of claim 1, wherein said set of polynucleotides further comprises at least one polynucleotide capable of hybridizing to mRNA encoded by a gene of unknown function.

5. The method of claim 1, wherein at least one of said first or second mammalian cell is further transduced with a polynucleotide to determine if said polynucleotide alters the expression of mRNA of said cell wherein said mRNA encodes a protein member of at least one of said biochemical pathways.

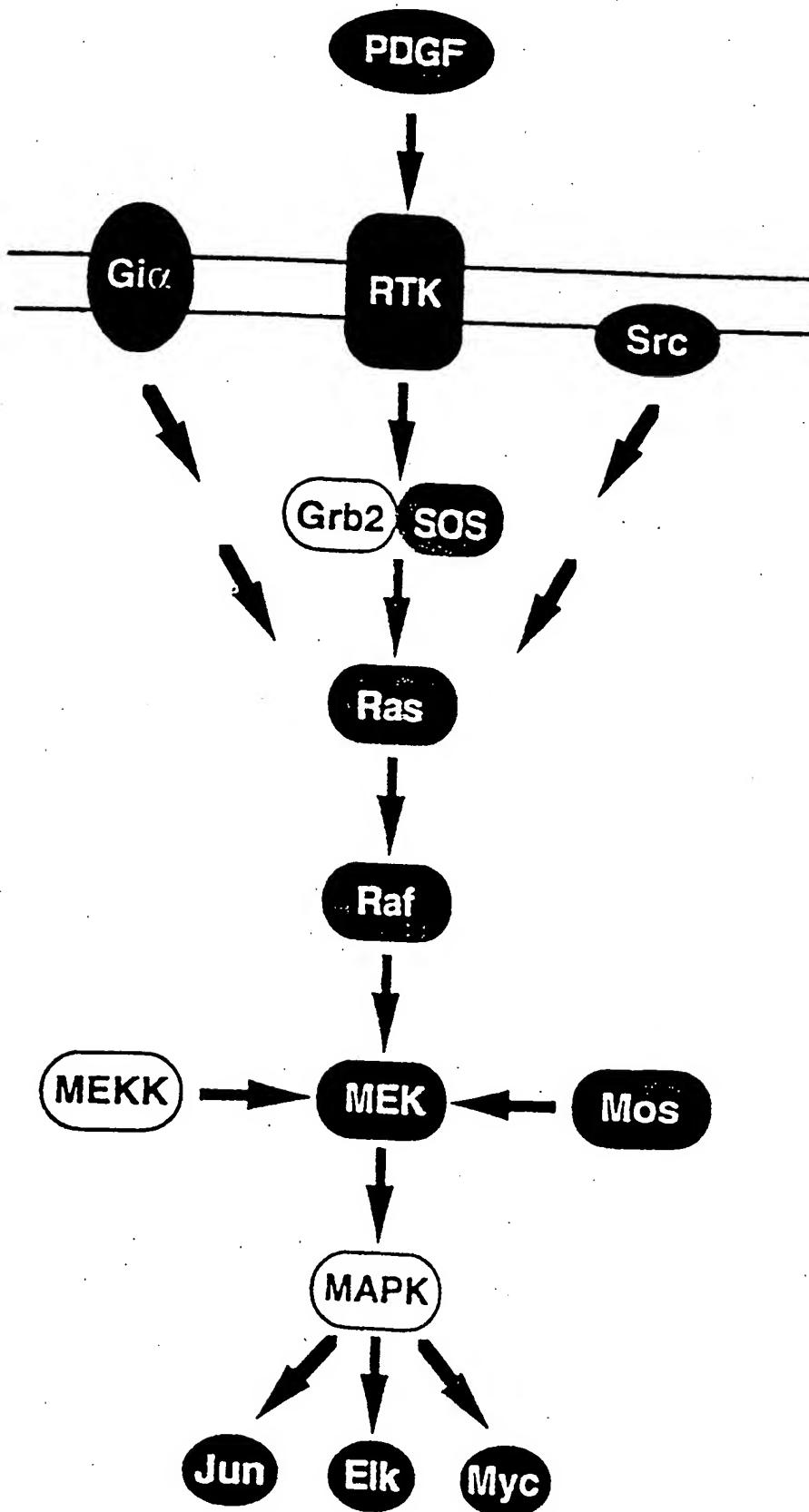


FIGURE 1

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/05260A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS, EMBASE, CHEM ABS Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 06277 A (UNIV CALIFORNIA) 20 February 1997 (1997-02-20) see whole doc. esp. claim 9 -----	1-5
X	WO 97 13877 A (LYNX THERAPEUTICS INC ;MARTIN DAVID W (US)) 17 April 1997 (1997-04-17) see whole doc. esp. claim 9 -----	1-5
X	WO 98 53103 A (BIBILASHVILLI ROBERT ;CHENCHIK ALEX (US); JOKHADZE GEORGE (US); CL) 26 November 1998 (1998-11-26) see whole doc. esp. p.149, 1.25 ff. and claims -----	1-5

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the International filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the International filing date but later than the priority date claimed

- "T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search  17 July 2000	Date of mailing of the international search report  21/07/2000
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Müller, F

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/05260

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9706277	A 20-02-1997	US 5569588 A		29-10-1996
		AU 6720996 A		05-03-1997
		CA 2202154 A		20-02-1997
		EP 0791078 A		27-08-1997
		JP 10507647 T		28-07-1998
WO 9713877	A 17-04-1997	AU 712929 B		18-11-1999
		AU 4277896 A		06-05-1996
		AU 6102096 A		30-12-1996
		AU 7717596 A		30-04-1997
		CN 1193357 A		16-09-1998
		CZ 9700866 A		17-09-1997
		CZ 9703926 A		17-06-1998
		EP 0793718 A		10-09-1997
		EP 0832287 A		01-04-1998
		EP 0931165 A		28-07-1999
		FI 971473 A		04-06-1997
		HU 9900910 A		28-07-1999
		JP 11507528 T		06-07-1999
		JP 10507357 T		21-07-1998
		NO 971644 A		02-06-1997
		NO 975744 A		05-02-1998
		PL 324000 A		27-04-1998
		WO 9641011 A		19-12-1996
		AU 718357 B		13-04-2000
WO 9853103	A 26-11-1998	US 5994076 A		30-11-1999
		AU 7593398 A		11-12-1998
		EP 0988398 A		29-03-2000
		US 6077673 A		20-06-2000

&lt;210&gt; 49

&lt;211&gt; 337

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;400&gt; 49

Met Ile Asp Arg Gln Arg Met Gly Leu Trp Ala Leu Ala Ile Leu Thr  
1 5 10 15

Leu Pro Met Tyr Leu Thr Val Thr Glu Gly Ser Lys Ser Ser Trp Gly  
20 25 30

Leu Glu Asn Glu Ala Leu Ile Val Arg Cys Pro Gln Arg Gly Arg Ser  
35 40 45

Thr Tyr Pro Val Glu Trp Tyr Tyr Ser Asp Thr Asn Glu Ser Ile Pro  
50 55 60

Thr Gln Lys Arg Asn Arg Ile Phe Val Ser Arg Asp Arg Leu Lys Phe  
65 70 75 80

Leu Pro Ala Arg Val Glu Asp Ser Gly Ile Tyr Ala Cys Val Ile Arg  
85 90 95

Ser Pro Asn Leu Asn Lys Thr Gly Tyr Leu Asn Val Thr Ile His Lys  
100 105 110

Lys Pro Pro Ser Cys Asn Ile Pro Asp Tyr Leu Met Tyr Ser Thr Val  
115 120 125

Arg Gly Ser Asp Lys Asn Phe Lys Ile Thr Cys Pro Thr Ile Asp Leu  
130 135 140

Tyr Asn Trp Thr Ala Pro Val Gln Trp Phe Lys Asn Cys Lys Ala Leu  
145 150 155 160

Gln Glu Pro Arg Phe Arg Ala His Arg Ser Tyr Leu Phe Ile Asp Asn  
165 170 175

Val Thr His Asp Asp Glu Gly Asp Tyr Thr Cys Gln Phe Thr His Ala  
180 185 190

Glu Asn Gly Thr Asn Tyr Ile Val Thr Ala Thr Arg Ser Phe Thr Val  
195 200 205

Glu Glu Lys Gly Phe Ser Met Phe Pro Val Ile Thr Asn Pro Pro Tyr  
210 215 220

Asn His Thr Met Glu Val Glu Ile Gly Lys Pro Ala Ser Ile Ala Cys  
225 230 235 240

Ser Ala Cys Phe Gly Lys Gly Ser His Phe Leu Ala Asp Val Leu Trp  
245 250 255

Gln Ile Asn Lys Thr Val Val Gly Asn Phe Gly Glu Ala Arg Ile Gln  
260 265 270

Glu Glu Glu Gly Arg Asn Glu Ser Ser Ser Asn Asp Met Asp Cys Leu  
275 280 285

Thr Ser Val Leu Arg Ile Thr Gly Val Thr Glu Lys Asp Leu Ser Leu  
290 295 300

Glu Tyr Asp Cys Leu Ala Leu Asn Leu His Gly Met Ile Arg His Thr  
305 310 315 320

Ile Arg Leu Arg Arg Lys Gln Pro Ser Lys Glu Cys Pro Ser His Ile  
325 330 335

Ala